CS 886: Multiagent Systems

Multiagent Learning

Kate Larson

Cheriton School of Computer Science
University of Waterloo

November 5, 2008

Outline

Introduction

- So far we have focused on computing optimal/equilibrium strategies
- Another approach: **learn** how to play a game
  - Play the game many times
  - Update your strategy based on experience
- Why?
  - Some aspect of the game may be unknown to you
  - Other agents may not be playing in equilibrium
  - Computing an optimal strategy is hard
  - Learning is what people do
  - ...

Introduction

- So far we have focused on computing optimal/equilibrium strategies
- Another approach: **learn** how to play a game
  - Play the game many times
  - Update your strategy based on experience
- Why?
  - Some aspect of the game may be unknown to you
  - Other agents may not be playing in equilibrium
  - Computing an optimal strategy is hard
  - Learning is what people do
  - ...

Introduction

- So far we have focused on computing optimal/equilibrium strategies
- Another approach: **learn** how to play a game
  - Play the game many times
  - Update your strategy based on experience
- Why?
  - Some aspect of the game may be unknown to you
  - Other agents may not be playing in equilibrium
  - Computing an optimal strategy is hard
  - Learning is what people do
  - ...

Challenges

- There are other agents in the environment
  - Dynamic environment (true in single agent settings)
  - What others are learning depend on what our agent is learning
    - Complex global behaviour of the system
  - Difficult to separate *learning* from *teaching*

	L	R
T	1,0	3,2
B	2,1	4,0

Challenges

- There are other agents in the environment
  - Dynamic environment (true in single agent settings)
  - What others are learning depend on what our agent is learning
    - Complex global behaviour of the system
  - Difficult to separate *learning* from *teaching*

	L	R
T	1,0	3,2
B	2,1	4,0

Challenges

- There are other agents in the environment
  - Dynamic environment (true in single agent settings)
  - What others are learning depend on what our agent is learning
    - Complex global behaviour of the system
  - Difficult to separate *learning* from *teaching*

	L	R
T	1,0	3,2
B	2,1	4,0

Challenges

- There are other agents in the environment
  - Dynamic environment (true in single agent settings)
  - What others are learning depend on what our agent is learning
    - Complex global behaviour of the system
  - Difficult to separate *learning* from *teaching*

	L	R
T	1,0	3,2
B	2,1	4,0

Goals of Multiagent Learning

Or *What is meant by successful learning?*

- No clear answer
- Descriptive Theories
- Prescriptive Theories

Goals of Multiagent Learning

Or *What is meant by successful learning?*

- No clear answer
- Descriptive Theories
- Prescriptive Theories

Repeated Games

Typically

- Agents play a normal-form game (the stage game)
- They see what happened (and get the payoffs)
- They play again
- ...

Can be repeated finitely or infinitely

- Extensive-form game with subgame-perfect equilibrium being repetition of some NE of the stage game
- Are there other equilibria?

Repeated Games

Typically

- Agents play a normal-form game (the stage game)
- They see what happened (and get the payoffs)
- They play again
- ...

Can be repeated finitely or infinitely

- Extensive-form game with subgame-perfect equilibrium being repetition of some NE of the stage game
- Are there other equilibria?

Repeated Games

Typically

- Agents play a normal-form game (the stage game)
- They see what happened (and get the payoffs)
- They play again
- ...

Can be repeated finitely or infinitely

- Extensive-form game with subgame-perfect equilibrium being repetition of some NE of the stage game
- Are there other equilibria?

Repeated Games

Typically

- Agents play a normal-form game (the stage game)
- They see what happened (and get the payoffs)
- They play again
- ...

Can be repeated finitely or infinitely

- Extensive-form game with subgame-perfect equilibrium being repetition of some NE of the stage game
- Are there other equilibria?

Finitely-repeated Prisoners' Dilemma

	C	D
C	2,2	0,3
D	3,0	1,1

- What will the agents do in the last round?
- What will the agents do in the second last round?
- ...
- What is the equilibrium?

Finitely-repeated Prisoners' Dilemma

	C	D
C	2,2	0,3
D	3,0	1,1

- What will the agents do in the last round?
- What will the agents do in the second last round?
- ...
- What is the equilibrium?

Finitely-repeated Prisoners' Dilemma

	C	D
C	2,2	0,3
D	3,0	1,1

- What will the agents do in the last round?
- What will the agents do in the second last round?
- ...
- What is the equilibrium?

Finitely-repeated Prisoners' Dilemma

	C	D
C	2,2	0,3
D	3,0	1,1

- What will the agents do in the last round?
- What will the agents do in the second last round?
- ...
- What is the equilibrium?

Finitely-repeated Prisoners' Dilemma

	C	D
C	2,2	0,3
D	3,0	1,1

- What will the agents do in the last round?
- What will the agents do in the second last round?
- . . .
- What is the equilibrium?

Infinitely repeated games

- Utility?
  - If you add up the utility over infinitely many rounds, then everyone gets infinity!
- Limit of *average* payoff:

$$\lim_{n\to\infty} \sum_{1\leq t\leq n} \frac{u(t)}{n}$$

- *Discounted* payoff:

$$\sum_t \delta^t u(t) \text{ for some } \delta,\ 0 < \delta < 1$$

Infinitely repeated games

- Utility?
  - If you add up the utility over infinitely many rounds, then everyone gets infinity!
- Limit of *average* payoff:

$$\lim_{n \to \infty} \sum_{1 \leq t \leq n} \frac{u(t)}{n}$$

- *Discounted* payoff:

$$\sum_t \delta^t u(t) \text{ for some } \delta,\ 0 < \delta < 1$$

Infinitely repeated Prisoners' Dilemma

	C	D
C	2,2	0,3
D	3,0	1,1

Tit-for-tat strategy:

- Cooperate in first round
- In every later round do the same thing that the other player did in the previous round

Trigger strategy:

- Cooperate as long as everyone cooperates
- Once an agent defects, defect *forever*

Folk Theorem: Any utility vector can be realized in NE if and only if it is *feasible* and *enforceable*.

Infinitely repeated Prisoners' Dilemma

	C	D
C	2,2	0,3
D	3,0	1,1

Tit-for-tat strategy:

- Cooperate in first round
- In every later round do the same thing that the other player did in the previous round

Trigger strategy:

- Cooperate as long as everyone cooperates
- Once an agent defects, defect *forever*

Folk Theorem: Any utility vector can be realized in NE if and only if it is *feasible* and *enforceable*.

Infinitely repeated Prisoners' Dilemma

	C	D
C	2,2	0,3
D	3,0	1,1

Tit-for-tat strategy:

- Cooperate in first round
- In every later round do the same thing that the other player did in the previous round

Trigger strategy:

- Cooperate as long as everyone cooperates
- Once an agent defects, defect *forever*

Folk Theorem: Any utility vector can be realized in NE if and only if it is *feasible* and *enforceable*.

Infinitely repeated Prisoners' Dilemma

	C	D
C	2,2	0,3
D	3,0	1,1

Tit-for-tat strategy:

- Cooperate in first round
- In every later round do the same thing that the other player did in the previous round

Trigger strategy:

- Cooperate as long as everyone cooperates
- Once an agent defects, defect *forever*

Folk Theorem: Any utility vector can be realized in NE if and only if it is *feasible* and *enforceable*.

Fictitious Play

Early and simply learning rule

- Initialize beliefs about opponent's strategy
- Repeat
  - Play a best-response to assessed strategy of opponent
  - Observe opponent's actual play and update beliefs accordingly

Note that agent is oblivious to the other agent's utilities.

Properties of Fictitious Play

Definition

An action profile a is in steady state if whenever a is played in round t then it is played in round $t + 1$.

Theorem

If a pure strategy profile is a strict NE of a stage game, then it is a steady state of fictitious play in the repeated game.

Theorem

If the empirical distribution of each agent's strategies converges in fictitious play then it converges to a Nash equilibrium.

Properties of Fictitious Play

Definition

An action profile a is in steady state if whenever a is played in round t then it is played in round $t + 1$.

Theorem

If a pure strategy profile is a strict NE of a stage game, then it is a steady state of fictitious play in the repeated game.

Theorem

If the empirical distribution of each agent's strategies converges in fictitious play then it converges to a Nash equilibrium.

Regret-based Learning

Regret:

$$R_i(a_i, t) = \frac{1}{t-1}\left[\sum_{1\leq t'\leq t-1} u_i(a_i, a_{-i,t'}) - u_i(a_{i,t'}, a_{-i,t'})\right]$$

An algorithm has *zero-regret* if or each a_i, the regret for a_i becomes non-positive as t goes to infinity (almost surely) against any opponents

Regret-based Learning

Regret:

$$R_i(a_i, t) = \frac{1}{t-1}\left[\sum_{1 \le t' \le t-1} u_i(a_i, a_{-i,t'}) - u_i(a_{i,t'}, a_{-i,t'})\right]$$

An algorithm has *zero-regret* if or each a_i, the regret for a_i becomes non-positive as t goes to infinity (almost surely) against any opponents

Regret-based Learning

- Regret matching:

$$\sigma_i^{t+1} = \frac{R^t(a_i)}{\sum_{a' \in A_i} R^t(a')}$$

- Regret matching has zero regret.
- If all players use regret matching, then play converges to the set of *weak correlated equilibria*
- Other types of regret-based learning have different properties

Regret-based Learning

- Regret matching:

$$\sigma_i^{t+1} = \frac{R^t(a_i)}{\sum_{a' \in A_i} R^t(a')}$$

- Regret matching has zero regret.
- If all players use regret matching, then play converges to the set of *weak correlated equilibria*
- Other types of regret-based learning have different properties

Regret-based Learning

- Regret matching:

$$\sigma_i^{t+1} = \frac{R^t(a_i)}{\sum_{a' \in A_i} R^t(a')}$$

- Regret matching has zero regret.
- If all players use regret matching, then play converges to the set of *weak correlated equilibria*
- Other types of regret-based learning have different properties

Regret-based Learning

- Regret matching:

$$\sigma_i^{t+1} = \frac{R^t(a_i)}{\sum_{a' \in A_i} R^t(a')}$$

- Regret matching has zero regret.
- If all players use regret matching, then play converges to the set of *weak correlated equilibria*
- Other types of regret-based learning have different properties

Targeted Learning

- Assume that there is a limited set of possible opponents
  - Try to do well against these

Example: is there a learning algorithm that

1. Learns to best-respond against any stationary opponent (one that always plays the same mixed strategy), and
2. Converges to a Nash equilibrium when playing against a copy of itself (self-play)?

Targeted Learning

- Assume that there is a limited set of possible opponents
    - Try to do well against these

Example: is there a learning algorithm that

1. Learns to best-respond against any stationary opponent (one that always plays the same mixed strategy), and
2. Converges to a Nash equilibrium when playing against a copy of itself (self-play)?

Stochastic Games

- Multiple states $S = \{S_1, \dots, S_m\}$
  - Each state, S_i is a normal form game
- After a round, random transition to another state
  - Transition probabilities depend on state and action taken
- Typically discount utilities over time

Note:

- 1-state stochastic game = (infinitely) repeated game
- 1-agent stochastic game = Markov Decision Process (MDP)

Stochastic Games

- Multiple states $S = \{S_1, \ldots, S_m\}$
  - Each state, S_i is a normal form game
- After a round, random transition to another state
  - Transition probabilities depend on state and action taken
- Typically discount utilities over time

Note:

- 1-state stochastic game = (infinitely) repeated game
- 1-agent stochastic game = Markov Decision Process (MDP)

Stationary Strategies

- A stationary strategy specifies a mixed strategy for each state
  - Strategy does not depend on history
  - For example, in a repeated game, stationary strategy = always playing the same mixed strategy
- An equilibrium in stationary strategies always exists [Fink 64]
- For 2-player zero-sum stochastic games one can use Shapley's algorithm to solve ($\sim$Value Iteration)

Stationary Strategies

- A stationary strategy specifies a mixed strategy for each state
  - Strategy does not depend on history
  - For example, in a repeated game, stationary strategy = always playing the same mixed strategy
- An equilibrium in stationary strategies always exists [Fink 64]
- For 2-player zero-sum stochastic games one can use Shapley's algorithm to solve ($\sim$Value Iteration)

Stationary Strategies

- A stationary strategy specifies a mixed strategy for each state
  - Strategy does not depend on history
  - For example, in a repeated game, stationary strategy = always playing the same mixed strategy
- An equilibrium in stationary strategies always exists [Fink 64]
- For 2-player zero-sum stochastic games one can use Shapley's algorithm to solve ($\sim$Value Iteration)